AD-A049 689 PRINCETON UNIV N J DEPT OF STATISTICS F/G 12/1
ESTIMATING THE INTENSITY OF A POISSON PROCESS.(U)
SEP 77 G S WATSON N00014-75-C-0453
UNCLASSIFIED TR-131-SER-2 NL

| OF |

AD A049689

END
DATE
FILMED
3-78
DDC

2 sc

ESTIMATING THE INTENSITY OF A POISSON PROCESS

See back page for 1473

by

G. S. Watson
Princeton University

Technical Report No. 131, Series 2
Department of Statistics
Princeton University
September 1977

Research sponsored in part by a contract with the Office of Naval Research, Contract No. N00014-75-C-0453, awarded to the Department of Statistics, Princeton University.

ESTIMATING THE INTENSITY OF A POISSON PROCESS

G. S. Watson
Princeton University

1. Introduction.

There is an immense and increasing literature on point processes, especially Poisson processes in which the intensity is a function $\lambda(t)$ of a scalar parameter t, usually time. Summaries are given in Cox and Lewis (1966), Lewis (1972), 1975). There are many motivations. Lewis has often been concerned with times of breakdown of computers. My own interest was aroused by two problems--the effects of the environment on the deaths per day in a large city and a discussion of wildcat oil strikes in Alberta. The latter was merely a numerical example in a paper by Clevenson and Zidek (1975) but it suggested that this topic would be appropriate for this Symposium. However, our discussion will be theoretical--we will speak of events, not deaths or strikes--and its aim is to show the simplicity and unity of a number of statistical and mathematical methods.

If the chances that an event occurs or does not occur in a smaller interval $(t,t+dt)$ are given by $\lambda(t)dt$, and $1 - \lambda(t)dt$, respectively, independently of what happened before t, then the number of events $N(t)$ in $(0,t)$ is said to follow a non-homogeneous Poisson process (N.H.P.P.). The numbers of events in non-overlapping intervals are independent and have Poisson distributions whose means are the integrals of $\lambda(t)$ over the relevant time intervals. Classical examples are deaths due to horse kicks in the Prussian Army and clicks of a Geiger Counter.

If the times of events are marked on a time axis, they will be densest, on the average, where $\lambda(t)$ is greatest. In this respect, $\lambda(t)$ is like a probability density. One can show that if n events occur in $(0,T)$, then the times of the events may be treated as a random sample from a probability distribution with density

$$f(t) = \lambda(t)/\int_0^T \lambda(t)\,dt \quad . \tag{1.1}$$

Given a set of data consisting of the successive times of events from an N.H.P.P., the statistician will wish to estimate the <u>function</u> $\lambda(t)$ and perhaps test some hypotheses about it. In some circumstances, the form of $\lambda(t)$ may be given and the problem reduced to estimating some parameters in the formula for $\lambda(t)$.

Given a random sample from some distribution, precisely the same problems arise--estimating its density or parameters in a formula for it.

In both problems, the data are often grouped to begin with or later for convenience--deaths per day, strikes per month. The mathematical difficulties of estimating $\lambda(t)$ are greatly reduced by dividing the time axis into intervals of equal length δ. We will start with such data. Clevenson and Zidek discussed very similar topics in continuous time and were essentially forced by difficulties to make discrete approximations later.

The problem of estimating a function $\lambda(t)$, $0 < t < T$ or the function λ_t, $t = 0,1,2$ N-1 is clearly quite distinct, given a finite amount of data, from the more usual problem of estimating a few parameters. We will use the total mean square error to assess the quality of the estimator, i.e., we will want an estimator $\hat{\lambda}_t$ that makes

$$E \sum_0^{N-1} (\hat{\lambda}_t - \lambda_t)^2 \tag{1.2}$$

small. We will further consider a measure of smoothness of the sequence $\hat{\lambda}_t$ and require it to be small too, on the grounds that the true sequence λ_t is not rough, namely

$$E \sum_0^{N-1} (\Delta^P \hat{\lambda}_t)^2 \tag{1.3}$$

where $\Delta^p \hat{\lambda}_t$ is the p-th difference of the $\hat{\lambda}_t$ sequence.

By combining these two measures of variability, we will have something to optimize. It will turn out that the "best" estimator depends upon the unknown sequence so that, in practice, one needs to put in one's intuition. While this is familiar in time series analysis, it comes as surprise to some people since such results are concealed in elementary statistics by assuming a parametric form. More familiar is our restriction to a class of estimators linear in the counts of events.

Sometimes it is realistic to assume that $\lambda(t)$ is itself a random function. Thus, we will have two cases--fixed and random functions. Since $\lambda(t)$ is often the signal that something is present, it is often referred to as the signal, rather than as the function.

Of course, for the statistical analyses of particular point processes, the N.H.P.P. model may not be appropriate. It is beyond our scope to consider the general problem. However, in Section 6, we briefly mention another model. In Section 7, some ideas for simpler and more exploratory analyses are suggested.

2. A Time Series Approach to the Fixed Signal Case.

Suppose we have the number of events in N successive intervals and that these are denoted by $n_0, n_1, \ldots, n_{N-1}$ with mean values $\lambda_0, \lambda_1, \ldots, \lambda_{N-1}$ which we may think of first as constants--a "fixed signal". Since they are integrals of $\lambda(t)$ over intervals of length δ, if $\lambda(t)$ is a smooth function, the λ_t's will behave regularly, and be smoother than the n_t's. Therefore, it is reasonable to consider a moving average estimator of λ_t given by

$$\hat{\lambda}_t = \sum_{-K}^{K} a_k n_{t-k} \quad . \tag{2.1}$$

To avoid negative estimates, the a_k would usually be taken as non-negative. Since a constant record ought to be equal to its estimate, one usually supposes $\Sigma a_k = 1$. Then the

a_k's form a discrete probability distribution. In the absence of other ideas, it would be symmetric around zero, its support, determined by K, would be large if λ_t is a very smooth function, small if it is not. It is fairly standard to call $\hat{\lambda}_t$ a <u>filtered</u> version of the n_t series, and the set of a_k's, a <u>filter</u>. It is clear that smoothing reduces the variance but increases the bias.

To find an optimal filter, one might ask for the filter which minimizes the average mean square error

$$E \sum_{0}^{N-1} (\hat{\lambda}_t - \lambda_t)^2 \tag{2.2}$$

where, for this section we regard the λ_t's as constants and we make no assumptions about the a_k's. However, one cannot calculate $\hat{\lambda}_t$ for t's such that (2.1) calls for n_t's with t outside 0, ..., N-1. To overcome this, we put the n_t and λ_t equally spaced around the circumference of a circle. It is then natural to identify n_N with n_0, n_{N+1} with n_1 etc., and n_{-1} with n_{N-1} etc. The functions n_t, λ_t, $\hat{\lambda}_t$ then are defined for all integers and they have period N. If we define

$$\omega_j = \frac{2\pi j}{N}, \quad j=0,1,\ldots,N-1 \tag{2.3}$$

$\exp i\omega_j t$ also has period N. Multiplying (2.1) by $\exp i\omega_j t$ and adding over t from 0 to N-1, we find

$$\hat{\lambda}(\omega_j) = A(\omega_j)n(\omega_j) \tag{2.4}$$

where

$$\hat{\lambda}(\omega_j) = \sum_{0}^{N-1} \hat{\lambda}_t \exp i\omega_j t,$$

$$A(\omega_j) = \sum_{-K}^{K} a_k \exp i\omega_j k, \tag{2.5}$$

$$n(\omega_j) = \sum_{0}^{N-1} n_t \exp i\omega_j t,$$

where we make no assumptions about K. Since

$$\sum_{t=0}^{N-1} \exp i\omega_j t = \begin{cases} N & \text{if } j=0 \\ 0 & \text{if } j=1,\ldots,N-1 \end{cases} \quad (2.6)$$

we can verify that

$$\hat{\lambda}_t = \frac{1}{N} \sum_0^{N-1} \hat{\lambda}(\omega_j) \exp -i\omega_j t \quad (2.7)$$

and

$$\sum_0^{N-1} \hat{\lambda}_t{}^2 = \frac{1}{N} \sum_0^{N-1} |\hat{\lambda}(\omega_j)|^2 \quad (2.8)$$

with similar formulae for any other function of period N.

Thus

$$\sum_0^{N-1} (\hat{\lambda}_t - \lambda_t)^2 = \frac{1}{N} \sum_0^{N-1} |\hat{\lambda}(\omega_j) - \lambda(\omega_j)|^2$$

$$= \frac{1}{N} \sum_0^{N-1} |A(\omega_j) n(\omega_j) - \lambda(\omega_j)|^2 . \quad (2.9)$$

Hence, the criterion (2.2) may be rewritten: find $A(\omega)$ so that

$$E \sum_0^{N-1} |A(\omega_j) n(\omega_j) - \lambda(\omega_j)|^2 = \min ! \quad (2.10)$$

The j-th term in the left-hand side (l.h.s.) of (2.10) is

$$|A(\omega_j)|^2 E|n(\omega_j)|^2 - A(\omega_j) En(\omega_j) \bar{\lambda}(\omega_j)$$

$$- \bar{A}(\omega_j) E\bar{n}(\omega_j) \lambda(\omega_j) + |\lambda(\omega_j)|^2 . \quad (2.11)$$

Now

$$En(\omega_j) = \sum_0^{N-1} E(n_j) \exp i\omega_j t = \lambda(\omega_j) , \quad (2.12)$$

$$E|n(\omega_j)|^2 = E\, n(\omega_j)\bar{n}(\omega_j),$$

$$= \sum \sum E(n_t n_{t'}) \exp i\omega_j(t-t'),$$

$$= \sum_0^{N-1} \lambda_t + |\lambda(\omega_j)|^2 , \qquad (2.13)$$

since the n_t are independent for distinct t, $En_t = \lambda_t$, and $E\, n_t^2 = \lambda_t + \lambda_t^2$. Now (2.11) may be rewritten as

$$E|n(\omega_j)|^2 \left| A(\omega_j) - \frac{En(\omega_j)\bar{\lambda}(\omega_j)}{E|n(\omega_j)|} \right|^2$$

$$+ |\lambda(\omega_j)|^2 - \frac{|En(\omega_j)\bar{\lambda}(\omega_j)|^2}{E|n(\omega_j)|^2} \qquad (2.14)$$

so that every term in the sum (2.10) is minimized if we choose

$$A(\omega_j) = \frac{|\lambda(\omega_j)|^2}{\sum_0^{N-1} \lambda_t + |\lambda(\omega_j)|^2} , \qquad (2.15)$$

giving (2.10) a minimum value of

$$\sum_{j=0}^{N-1} \frac{\sum_0^{N-1} \lambda_t}{\sum_0^{N-1} \lambda_t + |\lambda(\omega_j)|^2} |\lambda(\omega_j)|^2 . \qquad (2.16)$$

The optimal filter, defined by (2.15), has the following properties. We note that $0 \leq A(\omega_j) \leq 1$, all j. Since $A(\omega_j) = A(\omega_{-j})$ is real, $a_k = a_{-k}$ (i.e., the filter is symmetric about zero):

For

$$a_k = \frac{1}{N} \sum_{0}^{N-1} A(\omega_j) \exp -ik\omega_j,$$

$$= \frac{1}{N} \sum A(\omega_{-j}) \exp -ik\omega_j,$$

$$= \frac{1}{N} \sum A(\omega_{-j}) \exp -i(-k)\omega_{-j},$$

$$= a_{-k} .$$

Since $\sum a_k = A(0)$, (2.15) gives ($\sum \lambda_t = \lambda(0) > 0$)

$$\sum a_k = \frac{\lambda(0)^2}{\lambda(0) + \lambda(0)^2} < 1,$$

so that for <u>all</u> filters considered here, there is a downwards bias. Finally, since $A(\omega_j) = A(\omega_{-j}) = A(\omega_{N-j})$, it will be convenient to arrange that $N = 2K + 1$ or $K = (N-1)/2$.

If $\lambda_t = \lambda$, $t = 0,\ldots,N-1$, then $\lambda(\omega_j) = 0$ unless $j = 0$ when it is

$$\sum_{0}^{N-1} \lambda_t = N\lambda$$

so that

$$A(0) = \frac{(N\lambda)^2}{N\lambda + (N\lambda)^2}, \quad A(\omega_j) = 0, \quad j = 1,\ldots,N-1.$$

Hence, the a_k sequence is constant, a_N say, and equal to $A(0)/N$. Thus, $N\, a_N \uparrow 1$ as $N\lambda \to \infty$ so that the downward bias disappears.

By contrast, if $\lambda_t = \lambda$ for all t except t_0 when $\lambda_{t_0} = \lambda+\delta$ with δ very large, then

$$\lambda(\omega_j) = \begin{cases} N\lambda + \delta & \text{if } j=0 , \\ \delta \cdot \exp i\omega_j t_0 & \text{if } j\neq 0 , \end{cases} \qquad (2.17)$$

so that

$$A(\omega_j) = \begin{cases} \dfrac{(N\lambda + \delta)^2}{(N\lambda + \delta) + (N\lambda + \delta)^2} , & j = 0 \\[2ex] \dfrac{\delta^2}{(N\lambda + \delta) + \delta^2} , & j \neq 0 . \end{cases}$$

If $\delta > N\lambda$, $A(\omega_j) \simeq 1$ for all j and

$$a_k \simeq \begin{cases} 1 , & k = 0 , \\ 0 , & k \neq 0 . \end{cases}$$

Given the signal in the process, this is very reasonable--one would not want to smooth at all since only one λ_t is out of line. It is important to note that the optimal filter is unaffected by t_0. It will pick up this signal just as well wherever it is, i.e., the filter is <u>time-invariant</u> with respect to the position of the signal.

This is true in general since moving the signal on t_0 time units is the same here as multiplying $\lambda(\omega_j)$ by $\exp -i\omega t_0$ and $A(\omega_j)$ is unaffected since it depends only on the modulus of $\lambda(\omega_j)$.

Finally, the computation of the time domain filter coefficients given $A(\omega_j)$ can be done by using the Fast Fourier Transform to calculate

$$a_k = \frac{1}{N} \sum_{j=0}^{N-1} A(\omega_j) \exp -i\omega_j k . \qquad (2.18)$$

Equally, given some idea of λ_t, it can be used to calculate the $\lambda(\omega_j)$ which will be needed to obtain the approximate $A(\omega_j)$. Naturally, in practice, the ideal filter cannot be used for the estimation of the quantity it depends on. In recurring problems, or in the testing or detection problems, discussed in Section 4, this <u>is</u> possible.

The circular device used above is an often exploited mathematical trick to make the mathematics easy. It will only lead to trouble if the a_k's do not fall off to zero rapidly as $|k|$ increases. The two examples show when we may expect it to work.

3. The Case of a Random $\lambda(t)$.

It was suggested in Section 1 that in many problems $\lambda(t)$ should be taken as a random function. In this case to the expectation in (2.2) and so (2.10) should be added another expectation $\underset{\lambda}{E}$ over all $\lambda(t)$. The analysis is otherwise the same. Looking at (2.11), (2.12), and (2.13), we need to evaluate

$$\underset{\lambda}{E}\ |\lambda(\omega_j)|^2 \quad \text{and} \quad \underset{\lambda}{E} \sum \lambda_t \ .$$

If $\lambda(t)$ is a second-order stationary process in continuous time, then λ_t will be second-order stationary in discrete time. As we cannot with grouped data do more than study the λ_t process, we will only make assumptions about it. These are the following:

$$E\ \lambda_t = \lambda$$

$$\mathrm{Cov}(\lambda_t,\ \lambda_{t+\tau}) = \rho(\tau), \quad \sum_{\tau} |\rho(\tau)| < \infty$$

$$f(\phi) = \frac{1}{2\pi} \sum_{-\infty}^{\infty} \rho(\tau) \exp i\tau\phi,$$

$$\lambda_t - \lambda = \int_{-\pi}^{\pi} e^{i\phi t}\, dZ(\phi), \tag{3.1}$$

$$\rho(\tau) = \int_{-\pi}^{\pi} e^{i\phi\tau}\, f(\phi) d\phi,$$

$$\underset{\lambda}{E}\ dZ(\phi) = 0,$$

$$\mathop{E}_{\lambda}\, dZ(\phi)\overline{dZ(\phi')} = 0\ , \quad \phi \neq \phi',$$

$$\mathop{E}_{\lambda}\, |dZ(\phi)|^2 = f(\phi)d\phi\ .$$

These are the basic relations and quantities for such a process. $f(\phi)$ is called the (power) spectral density of the process. Readers unacquainted with this theory need only regard the integral formula for $\lambda_t - \lambda$ as an expression for this deviation from zero as a sum of sinusoidal oscillations with uncorrelated amplitudes whose average magnitudes vary with the frequency as described by $f(\phi)$, as seen in the last three expectations. The fact that the Fourier series for $f(\phi)$ has coefficients proportional to the covariances $\rho(\tau)$ will not be important here.

Then

$$\lambda(\omega_j) = \sum_0^{N-1} \lambda_t \exp i\omega_j t$$

$$= \lambda \sum_0^{N-1} \exp i\omega_j t + \int_{-\pi}^{\pi} \sum_0^{N-1} \exp(i(\phi+\omega_j)t)dZ(\phi)\ .$$

So

$$\lambda(0) = N\lambda + \int_{-\pi}^{\pi} \frac{1 - e^{iN\phi}}{1 - e^{i\phi}}\, dZ(\phi),$$

$$\lambda(\omega_j) = \int_{-\pi}^{\pi} \frac{1 - e^{iN(\phi+\omega_j)}}{1 - e^{i(\phi+\omega_j)}}\, dZ(\phi)\ , \quad j \neq 0\ .$$

Hence

$$\mathop{E}_{\lambda}\, \lambda(0) = E \sum \lambda_t = N\lambda$$

$$\mathop{E}_{\lambda}\, \lambda(\omega_j) = 0\ , \quad j \neq 0$$

$$\mathop{E}_{\lambda}|\lambda(0)|^2 = N^2\lambda^2 + \int_{-\pi}^{\pi}\left|\frac{1-e^{iN\phi}}{1-e^{i\phi}}\right|^2 f(\phi)d\phi, \tag{3.2}$$

$$\mathop{E}_{\lambda}|\lambda(\omega_j)|^2 = \int_{-\pi}^{\pi}\left|\frac{1-e^{iN(\phi+\omega_j)}}{1-e^{i(\phi+\omega_j)}}\right|^2 f(\phi)d\phi. \tag{3.3}$$

The multiplier of $f(\phi)$ when divided by $2\pi N$ in these last two integrals is called the Fejér kernel, i.e.,

$$g(\phi) = \frac{1}{2\pi N}\left|\frac{1-e^{iN\phi}}{1-e^{i\phi}}\right|^2 = \frac{1}{2\pi N}\left|\frac{e^{\frac{iN\phi}{2}} - e^{\frac{iN\phi}{2}}}{e^{\frac{i\phi}{2}} - e^{-\frac{i\phi}{2}}}\right|^2$$

$$= \frac{1}{2\pi N}\left|\frac{\sin\frac{N\phi}{2}}{\sin\frac{\phi}{2}}\right|^2 . \tag{3.4}$$

It may be shown that $g(\phi)$ has an integral from $-\pi$ to π of unity and that it resembles more and more as $N\to\infty$ a Dirac delta function with singularity at $\phi = 0$. The same is true of $g(\phi+\omega_j)$ except that the peak is moved to $-\omega_j$.

Thus, for large N, (3.2) and (3.3) become approximately

$$\left.\begin{aligned} E|\lambda(0)|^2 &\simeq N^2\lambda^2 + 2\pi N\, f(0) \doteq N^2\lambda^2 \\ E|\lambda(\omega_j)|^2 &\simeq 2\pi N\, f(\omega_j) \end{aligned}\right\} \tag{3.5}$$

so that the analogue of (2.14) here reads

$$(N\lambda + N^2\lambda^2)\left|A(0) - \frac{N^2\lambda^2}{N\lambda + N^2\lambda^2}\right|^2 + N^2\lambda^2 - \frac{N^2\lambda^2}{N\lambda + N^2\lambda^2}$$

and

$$(N\lambda + 2\pi N\, f(\omega_j)) \left| A(\omega_j) - \frac{2\pi N\, f(\omega_j)}{N\lambda + 2\pi N\, f(\omega_j)} \right|^2 + 2\pi N\, f(\omega_j) -$$

$$- \frac{2\pi N\, f(\omega_j)}{N\lambda + 2\pi N\, f(\omega_j)}$$

so that for large N

$$\left. \begin{aligned} A(0) &\simeq 1 , \\ A(\omega_j) &\simeq \frac{2\pi\, f(\omega_j)}{\lambda + 2\pi\, f(\omega_j)} \end{aligned} \right\} . \qquad (3.6)$$

Since $f(\phi)$ has period 2π and is symmetrical about 0, it is symmetrical about π. If the realizations of λ_t are smooth, on the average $f(\phi)$ will be larger near the origin and smaller away from it. Thus, the interpretation of this filter is exactly as before.

4. Invariant Testing.

Watson (1974) studied the problem of detecting a peak of unknown position in the intensity function. Both the situations, times of events given and counts in equal intervals given, were discussed. The first case was reduced to testing for uniformity on a circle. To make the correspondence to testing probability densities for uniformity clearer, the data were divided by T to fit on a circle of unit perimeter. If the times are then $x_1 . \ldots x_M$, the locally most powerful invariant tests of Beran (1968) have the form: reject when T_M is large where

$$T_M = \sum_1^{\infty} |c_m|^2 \frac{2}{M} \left| \sum_{j=1}^{M} e^{2\pi i m x_j} \right|^2 \qquad (4.1)$$

and the c_m are the Fourier coefficients of $g(x)$, a density with period unity. The family of alternatives is $(1 - \varepsilon) + \varepsilon g(x - \phi)$ where ϕ is any number in $(0,1)$ and ε is small. An alternative form of (4.1) is

$$T_M = \frac{1}{M} \int_0^1 [\Sigma\{g(x_i - \phi) - 1\}]^2 d\phi$$

$$= M \int_0^1 [\frac{1}{M} \Sigma g(x_i - \phi) - 1]^2 d\phi \qquad (4.2)$$

where

$$M^{-1} \sum_1^M g(x_i - \phi)$$

is a kernel-type estimator of the density of x, a test suggested by Watson (1967) on intuitive grounds. The test T_M is invariant with respect to the position of ϕ of the alternative density $g(x - \phi)$. In our case,

$$g(x - \phi) = \frac{\lambda(T(x - \phi))}{\int_0^T \lambda(t)dt)}$$

where $\lambda(\cdot)$ has been made to have period T. If only the counts in N equal intervals

$$(\frac{j}{N}, \frac{j+1}{N}) \quad .$$

for $j = 0, \ldots, N-1$ are given, the natural approximation to (4.1) and (4.2) is

$$T_M' = \sum_{m=1}^{\infty} |d_m|^2 \frac{2}{M} \left| \sum_0^{N-1} n_j \exp \frac{2\pi ij}{N} \right|^2 \qquad (4.3)$$

where

$$d_m = c_m \frac{N(1 - \exp -2\pi im/N)}{2\pi im}, \; m > 0 . \qquad (4.4)$$

Then (4.3) was shown to reduce to

$$T_M' = \sum_{m=1}^{N-1} \left\{ \sum_{\ell=0}^{\infty} |d_m + N\ell|^2 \right\} \frac{2}{M} \left| \sum n_j \exp \frac{2\pi imj}{N} \right|^2 \qquad (4.5)$$

because of aliasing. Experiments with a special case of T_M, called U_M^2 (Watson (1974)) with $N = 20$, showed that the null distributions of T_M and T_M' were practically the same.

Thus, for invariant testing, often little will be lost by grouping and the optimal statistic (then (4.5)) is easily computed using the F.F.T. In the notation of this paper, (4.5) is

$$T_M' = \sum_{m=1}^{N-1} \frac{|\lambda(\omega_m)|^2}{\Sigma\lambda_t} |n(\omega_m)|^2 \qquad (4.6)$$

$$\simeq \sum_{m=1}^{N-1} |A(\omega_m)|^2 \; |n(\omega_m)|^2 \qquad (4.7)$$

if $|\lambda(\omega_m)|^2/\Sigma\lambda_t$ is always small. This is so, for example, in the first special case of Section 2--a constant signal. The test in the second case, a short very strong signal, is based on

$$\sum_{m=1}^{N-1} |n(\omega_m)|^2$$

which is the discrete form of a test suggested in Watson (1967) for this situation.

Thus, the grouped form of invariant testing is very close to the grouped estimation problem of Section 2. Both benefit from being discussed in terms of finite Fourier transforms. If one ignored the test theory of Beran and just knew the results of Section 2, a natural test statistic, to test

λ_t = constant, λ say, would be

$$\sum_{t=0}^{N-1} (\hat{\lambda}_t - \lambda)^2 = \frac{1}{N} \sum_{0}^{N-1} |A(\omega_j)n(\omega_j) - \lambda'(\omega_j)|^2 \qquad (4.8)$$

where

$$\lambda'(\omega_j) = \sum_{0}^{N-1} \lambda \exp i\omega_j t = \begin{cases} N\lambda & \text{if } j = 0 \\ 0 & \text{if } j \neq 0 . \end{cases}$$

But

$$A(0)n(0) = \frac{|\lambda(0)|^2}{\sum_{1}^{N} \lambda_t + |\lambda(0)|^2} \sum n_t .$$

Thus, the first term in (4.8) merely tests the value chosen λ for the constant λ_t. The remaining terms test whether λ_t is constant so one would use only them for the test. But this is (4.7). In fact, $|\lambda(\omega_m)|^2/\Sigma\lambda_t$ small, all $m \neq 0$, really means that λ_t does not vary much and so defines a neighbourhood of λ_t = constant. Since Beran-type tests are only locally most powerful invariant, it is not surprising that the two tests coincide in this case.

Alternatives may be local, intermediate or distant. The last concept is defined in Watson (1974) from probability distributions on the circle and corresponds to cases like the second example at the end of Section 2. The optimal invariant tests are based on suprema, the associated distributions are hard to find and Fourier analysis seems to be of no avail. In the present case of (2.15), this suggests the intuitive test--is the greatest of $n_0, n_1, \ldots, n_{N-1}$ too large to have occurred by chance? Since the n_t's are independent Poissons with the same unknown mean on the null hypothesis, the test which is independent of the mean uses the distribution of the largest frequency in a multinomial of N equal cells when the total frequency Σn_t is kept fixed. This is not an easy problem. It has been studied by Dudley (1971). In general then distant tests may be expected to have the form:

$$\max_{t=0,\ldots,N-1} \hat{\lambda}(t) > C$$

and to be difficult to use because of distributional troubles.

5. Spline Estimation.

In Sections 2 and 3, we have looked only at average mean square error of the estimator $\hat{\lambda}_t$ of λ_t. A method which has a Bayesian type of motivation arises if we insist also that $\hat{\lambda}_t$ be a smooth function of t. We treat here the situation of Section 2. Thus, to the criterion (2.2) we wish to add

$$c^2 \sum_1^N (\Delta^2 \hat{\lambda}_t)^2 \quad .$$

From (2.7)

$$\Delta^2 \hat{\lambda}_t = \frac{1}{N} \sum \hat{\lambda}(\omega_j) \Delta^2 \exp -i\omega_j t$$

$$= \frac{1}{N} \sum \hat{\lambda}(\omega_j) \exp -i\omega_j t \,(\exp i\omega_j - 1)^2 \; .$$

By the analogue of (2.8)

$$\sum_1^N (\Delta^2 \hat{\lambda}_t)^2 = \frac{1}{N} \sum_1^N \left| \hat{\lambda}(\omega_j)(\exp i\omega_j - 1)^2 \right|^2 \; .$$

Hence, our criterion will now be: find $A(\omega_j)$ so that

$$E\sum_{1}^{N} |A(\omega_j)n(\omega_j) - \lambda(\omega_j)|^2 + c^2|A(\omega_j)n(\omega_j)(\exp i\omega_j - 1)^2|^2$$

$$= \min! \tag{5.1}$$

In (5.1), c^2 is a smoothing parameter. As $c^2 \to \infty$, $\hat{\lambda}(t)$ must become flat.

Evaluating (5.1) by making use of the results that lead up to (2.13), we find that the j-th term in (5.1) is

$$|A(\omega_j)|^2(\sum\lambda_t + |\lambda(\omega_j)|^2)(1 + c^2|\exp i\omega_j - 1|^4)$$

$$- (A(\omega_j) + \bar{A}(\omega_j))|\lambda(\omega_j)|^2 + |\lambda(\omega_j)|^2$$

$$= (\sum\lambda_i + |\lambda(\omega_j)|^2)(1 + c^2|e^{i\omega_j} - 1|^4)\left|A(\omega_j) - \frac{|\lambda(\omega_j)|^2}{(\sum\lambda_i + |\lambda(\omega_j)|^2)(1 + c^2|e^{i\omega_j} - 1|^4)}\right|^2$$

$$+ \left\{|\lambda(\omega_j)|^2 - \frac{|\lambda(\omega_j)|^4}{(\sum\lambda_i + |\lambda(\omega_j)|^2)(1 + c^2|e^{i\omega_j} - 1|^4)}\right\} \tag{5.2}$$

where the last term in (5.2) is positive. Thus, the optimal filter now is given by

$$A(\omega_j) = \frac{|\lambda(\omega_j)|^2}{(\sum\lambda_t + |\lambda(\omega_j)|^2)(1 + c^2|\exp i\omega_j - 1|^4)} . \tag{5.3}$$

Clearly, the result (5.3) reduces to (2.15) when $c^2 = 0$. When $c^2 \to \infty$, $A(\omega_j) \to 0$ unless $j = 0$ so that the a_k become equal and the estimator becomes constant for $t = 1, \ldots, N$ as predicted above.

As in Section 2, it is easy to calculate the time filter coefficients a_k using the Fast Fourier Transform.

The use of

$$\sum_1^N (\Delta^2 \hat{\lambda}_t)^2$$

to control smoothness is common but had we chosen

$$\sum_1^N (\Delta^p \hat{\lambda}_t)^2$$

the optimal filter would have been

$$A(\omega_j) = \frac{|\lambda(\omega_j)|^2}{(\sum \lambda_t + |\lambda(\omega_j)|^2)(1 + c^2 |\exp i\omega_j - 1|^{2p})} \tag{5.4}$$

and its properties for $p \geq 1$ are the same.

Splines in the discrete case were first suggested by Whittaker (1923) and in the continuous case by Schoenberg (1964). To obtain a Bayesian motivation for (5.1), it is usual to invent a process for λ_t which makes the likelihood function depend on $\sum (\Delta^p \lambda_t)^2$. This would be so if $\Delta^p \lambda_t$ for the various values of t were independently Gaussian with the same variance. In continuous time, this requires $\lambda(t)$ to be an integrated Brownian motion.

6. Another Model.

Given the above development, it seems worthwhile to show how the Weiner Filter may be easily derived. This means changing the model from the N.H.P.P. In fact, we have only assumed, in the Section 2 derivation of the optimal linear filter, that

$$E(n_t) = \lambda_t , \text{var}(n_t) = \lambda_t , \text{cov}(n_t , n_{t'}) = 0, \; t \neq t' . \tag{6.1}$$

Instead, suppose that

$$E(n_t) = \lambda_t \text{ , } \operatorname{cov}(n_t \text{ , } n_{t+\tau}) = \rho(\tau) \text{ .} \qquad (6.2)$$

Thus, the counts in different intervals may be correlated but this correlation should depend only upon their spacing. In this sense, the assumptions are weaker. However, $\rho(0) = \operatorname{var}(n_t)$ must now be constant and not depend upon λ_t.

Then (2.13) becomes

$$E|n(\omega_j)|^2 = \sum\sum E(n_t n_{t'}) \quad \exp i\omega_j(t-t')$$

$$= |\lambda(\omega_j)|^2 + \sum\sum\rho(t-t') \exp i\omega_j(t-t'). \qquad (6.3)$$

Given the fact that we have extended the data to make n_t of period N, $\rho(\tau)$ in (6.2) must have period N. Thus, $\rho(-1) = \rho(N-1)$, $\rho(-2) = \rho(N-2)$, etc., so that the covariance matrix C whose element in the t-th row, t'-th column is $\rho(t-t')$ is an N×N circulant. The eigenvectors of such matrices are $[\exp i\omega_j \cdot 0, \exp i\omega_j \cdot 1, \ldots, \exp i\omega_j(N-1)]$, $j = 0, \ldots, N-1$. Denoting the corresponding eigenvalues by $f(\omega_j)$, (6.3) becomes

$$E|n(\omega_j)|^2 = |\lambda(\omega_j)|^2 + f(\omega_j) \text{ .} \qquad (6.4)$$

In this case, the analogue of (2.15) is seen to be

$$A(\omega_j) = \frac{|\lambda(\omega_j)|^2}{|\lambda(\omega_j)|^2 + f(\omega_j)} \text{ .} \qquad (6.5)$$

This is nothing but the classical result of Wiener for filtering a stationary process--except that we have derived it on the circle.

In order to use (6.5) in practice to derive a good smoothing formula, one needs to make a good guess at the "noise" spectrum $f(\omega_j)$ as well as the signal spectrum $|\lambda(\omega_j)|^2$, which was all that was needed in Section 2. This is the price of the weaker assumption.

7. An Even More Applied Approach.

A data analyst, knowing that $E(n_t) = \text{Var}\,(n_t) = \lambda_t$, would probably consider $y_t = 2\sqrt{n_t}$ which will have approximately unit variance and mean $2\sqrt{\lambda_t}$, provided λ_t is not small. In passing a curve through the points (t, y_t), the analyst will not be distracted by variance changes. Further y_t will be approximately Gaussian so it makes sense to pass the curve through the "center" of the band of points. Finally, any smoothing method could be considered, without regard to positivity, since, given the smoothed values, $\hat{y}_t$ say, a positive smoothed estimate of λ_t is available from

$$\hat{\lambda}_t = \frac{1}{4}\hat{y}_t^{\,2} \; . \tag{7.1}$$

If other transformations such as $2\sqrt{n_t + 3/8}$ or $\sqrt{n_t} + \sqrt{n_t+1}$ are used, the formulae replacing (7.1) are obtained similarly.

To counter criticisms, this analyst might object to (1.2) as a criterion on the grounds that the derivations $\hat{\lambda}(t) - \lambda(t)$ are weighted equally despite the fact that $\text{Var}(\hat{\lambda}(t) - \lambda(t)) = O(\lambda(t))$. The analyst might also criticize the notion of optimal smoothing by saying that, since the true $\lambda(t)$ is not known, one must try various smoothers on the data, possibly smoothing less in some stretches than in others to retain resolution.

It is, of course, possible to formalize smoothing with y_t, the square roots of the counts. A standard procedure would be to use a spline fit, e.g., find $\hat{\mu}_t = 2\sqrt{\lambda_t}$ to minimize

$$\sum (y_t - \hat{\mu}_t)^2 + K \sum (\Delta^p \hat{\mu}_t)^2 \; .$$

The first term is more justifiable now since the y_t have approximately the same variance about μ_t. The second term insists on smoothness of the fitted μ_t's. The K is a tuning parameter, $K = 0$ implying no smoothing at all, i.e., $\hat{\mu}_t = y_t$.

References.

Beran, R. J. W. (1968). "Testing for uniformity on a compact homogeneous space," *J. Appl. Prob.* *5*, pp. 177-195.

Bloomfield, P. (1976). *Fourier Analysis of Time Series - An Introduction*. Wiley-Interscience, New York.

Brillinger, D. R. (1974). *Time Series, Data Analysis and Theory*. Holt, Rinehart and Winston, New York.

Clevenson, M. L., and Zidek, J. V. (1975). "Bayesian estimation of the intensity function of the non-stationary Poisson process," Technical Report No. 222, Department of Statistics, Stanford University.

Cox, D. R., and Lewis, P. A. P. (1966). *The Statistical Analysis of Series of Events*. Methuen, London.

Dudley, R. M. (1971). "Binomial tests of goodness of fit," Unpublished Manuscript.

Grandell, J. (1971). "On stochastic processes generated by a stochastic intensity," *Skand. Akt. Tid.* *54*, pp. 205-240.

Lewis, P. A. P. (1972). *Stochastic Point Processes: Statistical Analysis and Applications*. Wiley-Interscience, New York.

Lewis, P. A. P. (1975). "Statistical analysis of transacting processing in a data base system," IBM Report No. RJ1629.

Lundberg, O. (1940). *On Random Processes and Their Applications to Sickness and Accident Statistics*. Almquist and Wiksells, Uppsala. (Republished 1964).

Schoenberg, I. J. (1964). "Spline functions and the problem of graduation," *P. N. A. S.* *52* (4), pp. 947-950.

Watson, G. S. (1961). "Goodness of fit tests on a circle," *Biometrika* *48*, pp. 109-114.

Watson, G. S. (1967). "Some problems in statistics of directions," *Bull. I. S. I.*, 33rd Session, pp. 374-385.

Watson, G. S. (1974a). "Detecting signals in Poisson noise," in *Reliability and Biometry*, pp. 775-786, ed., F. Proschan and R. J. Serfling, S. I. A. M., Philadelphia.

Watson, G. S. (1974b). "Optimal invariant tests for uniformity," *Studies in Probability and Statistics*. Jerusalem Academic Press, pp. 121-128.

Whittaker, E. T. (1923). "On a new method of graduation," *Proc. Edin. Math. Soc.* *41*, pp. 63-75.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

14 TR-131-SER-2

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report No. 131	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) 6 ESTIMATING THE INTENSITY OF A POISSON PROCESS.		5. TYPE OF REPORT & PERIOD COVERED 9 Technical rept.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) 10 G. S. Watson		8. CONTRACT OR GRANT NUMBER(s) 15 N00014-75-C-0453
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Princeton University Princeton, N. J. 08540		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research (Code 436) Arlington, Va. 22217		12. REPORT DATE 11 Sep[redacted]77 12 24p.
		13. NUMBER OF PAGES 22
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) non-homogeneous Poisson process, smoothing, splines, intensity estimation, moving averages		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) A linear estimator minimizing the expected average mean square error is derived. This is generalized by adding to the criterion a multiple of the expected average squared p^{th} difference. This leads to splines.		

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED 406 873

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)